

PENERAPAN *MACHINE LEARNING* DALAM MELAKUKAN PREDIKSI HARGA SAHAM PT. BANK MANDIRI (PERSERO) TBK DENGAN ALGORITMA *LINEAR REGRESSION*

Pita Triya, Nana Suarna, Nisa Dienwati Nuris

Komputerisasi Akuntansi, STMIK IKMI Cirebon

Jalan Perjuangan No.10B Karyamulya Kesambi Kota Cirebon, Indonesia

pitatriya09@gmail.com

ABSTRAK

Investor dalam melakukan kegiatan investasi saham tentunya mengharapkan return saham yang positif sesuai yang diharapkan. Investor harus melakukan analisis sebelum membeli saham dengan menggunakan pendekatan fundamental dan teknikal karena investasi saham dikenal dengan istilah *high gain high risk*. Analisis teknikal melibatkan evaluasi pergerakan harga saham pada periode sebelumnya. Untuk meraih keuntungan, perubahan harga saham perlu diamati oleh investor secara *time series* dan resiko yang bisa terjadi. *Machine Learning* telah merepresentasikan keunggulan yang luar biasa dalam melakukan prediksi pasar saham. Metode dalam penelitian menggunakan algoritma regresi linier dengan menggunakan aplikasi Rapidminer. Algoritma regresi linier ini dipilih karena mampu membuat suatu estimasi dengan memanfaatkan data-data lama, sehingga dapat dihasilkan suatu pola hubungan antara atribut-atribut yang mempengaruhi hasil prediksi. Penelitian ini bertujuan untuk membuat model *machine learning* yang dapat memprediksi harga saham PT. Bank Mandiri (Persero) Tbk berdasarkan data *historis* saham menggunakan 6 variabel. Dengan 5 variabel independen yaitu *Date*, *Open*, *High*, *Low*, *Volume*, dan variabel dependen *Close*. Data yang digunakan selama 5 tahun dari tanggal 01 november 2018 sampai 29 november 2023 yang diperoleh dari website *finance.yahoo.com*. menghasilkan nilai *root mean squared error* terbaik 0.005, *absolute error* 0.004, *relative error* 2.04%, *correlation* 0.999, dan *squared correlation* 0.998.

Kata Kunci : *Stock Prices, Predictions, Machine Learning, Linear Regression*

1. PENDAHULUAN

Investasi saham menjadi salah satu kegiatan yang sedang populer selama beberapa tahun belakangan. Berdasarkan data dari Kustodian Efek Indonesia (KSEI) jumlah investor saham di Indonesia pada tahun 2022 yaitu 4.439.933 yang mengalami kenaikan sebesar 28,64% dari tahun sebelum nya yaitu pada tahun 2021 jumlah investor saham berjumlah 3.451.513. Hal ini menggambarkan bahwa banyak masyarakat Indonesia sudah mulai sadar untuk melakukan investasi saham. Perkembangan yang sangat signifikan ini juga didukung dengan banyak nya aplikasi yang dikeluarkan oleh broker (perusahaan efek) untuk melakukan kegiatan jual beli saham secara praktis.

Saham merupakan tanda kepemilikan hak atas suatu perusahaan dalam bentuk surat berharga [1]. Investasi saham memiliki istilah *high gain high risk* dimana selain memberikan keuntungan yang tinggi, juga dapat memberikan resiko yang tinggi. Resiko tinggi yang muncul pada investasi menjadi pertimbangan para investor untuk melakukan analisis sebelum membeli saham. Terdapat dua analisis yang

umumnya digunakan dalam investasi saham yaitu analisis fundamental dan analisis teknikal. Analisis fundamental adalah analisis yang dilakukan berdasarkan pada kondisi perusahaan di masa terdahulu dan kemungkinan yang terjadi di masa mendatang dalam kaitan untuk memprediksi harga saham dengan menggunakan informasi dari berita ataupun laporan keuangan perusahaan. Kemudian, analisa teknikal dilakukan berdasarkan pada pergerakan harga saham di periode sebelumnya. *Machine Learning* merupakan metode akurat yang dapat digunakan dalam melakukan prediksi harga saham [2]. *Linear Regression* merupakan model regresi dengan teknik untuk menganalisis estimasi nilai variabel dependen 'y' dengan rentang nilai variabel independen 'x'. Regresi linier berganda merupakan teknik statistika untuk memprediksi hasil dari variabel jawaban dengan menggunakan sejumlah variabel penjelas [3].

Tabel 1. Data historis saham dari website yahoo.finance

No.	Date	Open	High	Low	Close	Adj Close	Volume
1	01/11/2018	3.450	3.675	3.438	3.613	2.925	185151800
2	02/11/2018	3.663	3.750	3.550	3.713	3.006	143785200
3	05/11/2018	3.700	3.700	3.625	3.688	2.986	56013400
4	06/11/2018	3.738	3.738	3.625	3.675	2.976	124000800
5	07/11/2018	3.688	3.700	3.600	3.675	2.976	92281000

No.	Date	Open	High	Low	Close	Adj Close	Volume
6	08/11/2018	3.700	3.763	3.688	3.750	3.037	130444200
7	09/11/2018	3.625	3.663	3.575	3.613	2.925	123465600
8	12/11/2018	3.613	3.613	3.488	3.488	2.824	65689800
9	13/11/2018	3.400	3.588	3.400	3.575	2.895	85750000
10	14/11/2018	3.638	3.650	3.563	3.563	2.885	68203400
11	15/11/2018	3.625	3.725	3.575	3.700	2.996	115193000
12	16/11/2018	3.738	3.813	3.688	3.725	3.016	113146000
...
1254	28/11/2023	5.900	5.925	5.875	5.900	5.900	58255800
1255	29/11/2023	5.850	5.925	5.825	5.850	5.850	36796600

Berdasarkan tabel 1. merupakan data historis harga saham dari PT. Bank Mandiri Tbk, selama 5 tahun, dari tanggal 1 november 2018 sampai 29 november 2023 dengan jumlah data sebanyak 1255 baris data public yang didapat dari website resmi *finance.yahoo.com*. Berdasarkan data historis saham terdapat 7 atribut yaitu *date*, *open*, *high*, *low*, *close*, *adj close* dan *volume*. Atribut *date* berupa tanggal, bulan dan tahun perdagangan pasar saham. *Open* merupakan harga awal atau harga pembuka pada perdagangan pasar saham dihari tersebut. *High* merupakan harga saham tertinggi yang dapat dicapai pada perdagangan pasar saham. *Low* merupakan harga saham terendah yang dapat dicapai pada perdagangan pasar saham. *Close* merupakan harga penutupan perdagangan pasar saham yang berakhir dihari tersebut. *Adj Close* merupakan harga penutupan perdagangan pasar saham yang sudah disesuaikan ketika terjadi aksi korporasi perusahaan. *Volume* merupakan jumlah saham yang diperdagangkan pada periode tertentu [4].

Dilansir dari website resmi *idxchannel.com* saham bank mandiri tergolong dalam indeks LQ45 untuk periode Agustus 2022 hingga Januari 2023. Saham yang masuk dalam indeks LQ45 dipilih berdasarkan kriteria likuiditas dan nilai kapitalisasi pasar, serta didukung oleh fundamental perusahaan [5]. Menurut informasi dari situs resmi Bursa Efek Indonesia (IDX), Bank Mandiri mencatat kinerja yang mengungguli industri perbankan Indonesia pada tahun 2022. Pada akhir September 2023, Bank Mandiri mencapai pencapaian luar biasa sebagai bank pertama di Indonesia dengan total aset konsolidasi mencapai Rp 2.007 triliun, mengalami pertumbuhan sekitar 9,11% dari tahun sebelumnya(2022). Kenaikan ini sebagian besar dipicu oleh pertumbuhan kredit dari sumber pihak ketiga yang berhasil tumbuh positif. Informasi ini tentu menarik perhatian investor atau calon investor yang ingin mengetahui prospek harga saham perusahaan ini di masa mendatang. Untuk memperoleh prediksi harga saham yang lebih akurat, salah satu pendekatan yang dapat digunakan oleh investor adalah dengan memanfaatkan teknologi *machine learning*.

2. TINJAUAN PUSTAKA

2.1. Literature Review

Penelitian sebelumnya oleh Dwi Ismiyana Putri dan Mardi Yudhi Putra dalam Jurnal Khatulistiwa Informatika pada tahun 2023 dengan judul

“Komparasi Algoritma Dalam Memprediksi Perubahan Harga Saham Goto Menggunakan RapidMiner”. Penelitian ini menggunakan atribut *timestamp*, *open*, *low*, *high*, *close*, *volume* dengan variabel *close* sebagai label dalam dataset. Hasil analisa yang dilakukan penelitian tersebut menemukan bahwa model *Neural Network* (NN) memberikan hasil RMSE yang paling rendah yaitu sebesar 16.562. Dapat disimpulkan bahwa model *Neural Network* memiliki tingkat akurasi prediksi lebih baik dibandingkan dengan model Regresi Linier sebesar 22.479 dan Support Vector Machine sebesar 31.326 dalam memprediksi perubahan harga saham GOTO [6].

Penelitian yang dilakukan oleh Janur Syahputra, Rima Dias Ramadhani, dan Auliya Burhanudin dalam jurnal Institut Teknologi Telkom Purwokerto pada tahun 2022 dengan judul “Prediksi Harga Saham Bank BRI Menggunakan Algoritma Linear Regression Sebagai Strategi Jual Beli Saham”. Variabel yang digunakan yaitu *date* dan *price*. Penelitian tersebut menghasilkan rasio terbaiknya di angka 80:20 dengan menghasilkan akurasi train dan test sebesar 0.89 dan 0.91. Hasil eror dari prediksi dihitung menggunakan MAPE dan menghasilkan persentase sebesar 13.751% sebagai data pelatihan, 13.773% sebagai data pengujian, dan 13,755% sebagai data keseluruhan. Dari hasil MAPE tersebut dapat disimpulkan bahwa metode *linear regression* dapat digunakan untuk melakukan prediksi harga saham Bank BRI [7].

Penelitian yang dilakukan oleh Merfin dan Raymond Sunardi Oetama dalam Jurnal Sistem Informasi pada tahun 2019 dengan judul “Prediksi Prospek Harga Saham Perusahaan Perbankan Menggunakan Regresi Linier (Studi Kasus Bank BCA Tahun 2015-2017)”. Atribut yang digunakan dalam penelitian tersebut yaitu *Open*, *High*, *Low*, *Close*, *AdjClose*, serta *Volume* dan menambahkan atribut *closing*. Hasil dari penelitian tersebut yaitu hasil MAPE pada bulan April tahun 2015 menjadi yang terbesar dengan angka MAPE sebesar 14.765,67. Untuk hasil MAPE terkecil dengan angka 234,64 terdapat pada bulan Mei tahun 2017. Rekomendasi yang tepat untuk hasil evaluasi MAPE tersebut adalah tahun 2016 merupakan tahun dimana masyarakat memulai investasi saham [8].

2.2. Saham

Saham merupakan dokumen kepemilikan seseorang atau entitas terhadap saham suatu

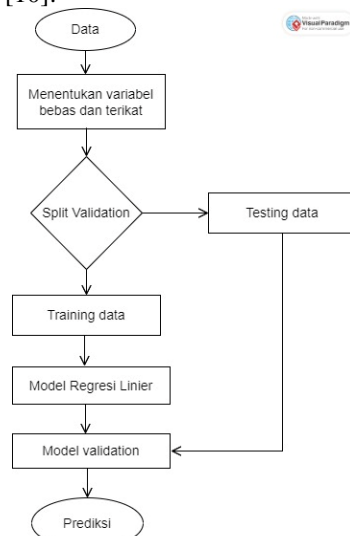
perusahaan. Perusahaan memiliki berbagai opsi untuk mendapatkan sumber pendanaan, baik dari internal perusahaan maupun eksternal perusahaan. Sumber pendanaan internal umumnya berasal dari laba yang ditahan. Sementara itu, alternatif pendanaan eksternal dapat melibatkan penerbitan saham atau pengambilan utang dari kreditur, juga melalui penawaran obligasi [5].

2.3. Prediksi

Prediksi merupakan kegiatan yang bertujuan untuk meramalkan kejadian yang mungkin terjadi di masa depan. Proses peramalan ini melibatkan pengukuran berdasarkan analisis ilmiah menggunakan metode statistika dengan maksud untuk meningkatkan ketepatan dalam meramalkan peristiwa di masa mendatang. Secara sederhana, tujuan dari prediksi ini adalah untuk menghasilkan estimasi yang dapat mengurangi kesalahan dalam peramalan (forecast error), yang dapat diukur dengan menggunakan *Standart Error Estimate* (SEE), *Mean Absolute Percent Error* (MAPE), *Absolute Error*, *Relative Error*, *Roor Mean Squared Error* (RMSE) dan alat ukur lainnya [9].

2.4. Regresi Linier

Regresi Linier memungkinkan dalam hal pemodelan hubungan antar variabel independen (input) dengan variabel dependen (output) dalam bentuk persamaan linier. Dengan menggunakan regresi linier kita dapat memprediksi atau mengestimasi nilai variabel dependen berdasarkan nilai variabel independen yang ada. Dalam konteks data science, regresi linier menjadi salah satu alat yang penting untuk memahami dan menggali pola, serta hubungan antara variabel yang terlibat dalam suatu dataset. Metode regresi linier membantu kita untuk memahami sejauh mana suatu variabel independen berkontribusi terhadap variabel-variabel dependen, serta melihat bagaimana perubahan nilai variabel independen dapat memengaruhi nilai variabel dependen [10].



Gambar 1. Flowchart regresi linier

Berdasarkan gambar 1, berikut merupakan tahapan metode regresi linier:

- Tentukan terlebih dahulu variabel apa saja yang akan dijadikan variabel bebas dan terikat.
- Lakukan *split validation* untuk membagi data menjadi data *training* atau data latih dan data *testing* atau data uji. Rasio perbandingan data yang biasa digunakan yaitu rasio jumlah data *training* harus lebih banyak dibanding data *testing*.
- Data latih akan digunakan oleh algoritma regresi linier untuk dilakukan dalam membangun model.
- Dilakukan proses algoritma regresi linier untuk menemukan pola hubungan antar variabel.
- Pada tahap model *validation*, data *testing* akan digunakan menjadi pengujian untuk membandingkan prediksi model yang sudah dibuat.
- Menghasilkan nilai *koefisien* dan *intercept* dari model yang dibangun untuk melakukan prediksi.

Berikut persamaan matematik regresi linier berganda:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

Dimana:

Y = variabel tak bebas (nilai variabel yang akan diprediksi)

β_0 = *intercept* atau nilai konstan dalam model

$\beta_1, \beta_2, \dots, \beta_n$ = nilai koefisien regresi yang menggambarkan hubungan antara Y dengan X_1

X_1, X_2, \dots, X_n = variabel bebas atau *variabel prediktor*

Serta menghasilkan nilai error RMSE, dengan rumus:

$$RMSE = \sqrt{\frac{\sum (y_k - \hat{y}_k)^2}{N}} \quad (2)$$

Dimana:

y_k : nilai asli

\hat{y}_k : nilai hasil prediksi

N : jumlah sampel data

- Menghasilkan *absolute error* dengan persamaan:

$$AE = AV - MV \quad (3)$$

Dimana:

AV = *Actual Value* (nilai aktual)

MV = *Measured Value* (nilai rata-rata)

- Menghasilkan *relative error* dengan persamaan:

$$RE = \frac{AV - MV}{AV} \times 100 \quad (4)$$

Dimana:

AV : *Actual Value* (nilai aktual)

MV : *Measured value* (nilai rata-rata)

- Menghasilkan *correlation* atau koefisien korelasi dengan persamaan:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}} \quad (5)$$

Dimana:

r = koefisien korelasi

x = nilai variabel bebas

y = nilai variabel terikat

\bar{y} = rata-rata variabel terikat

\bar{x} = rata-rata variabel bebas

- Menghasilkan *squared correlation* atau Koefisien Determinasi dengan persamaan:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

Dimana:

R^2 = Koefisien determinasi

y_i = Variabel terikat pengamatan i

\hat{y}_i = Nilai yang didekati dengan model regresi untuk observasi i

\bar{y} = rata-rata variabel terikat di semua pengamatan

2.5. Data Mining

Data mining adalah bidang ilmu yang berorientasi dalam pengungkapan pola dan pengetahuan yang tersembunyi pada sebuah data yang besar. Bidang ini berada pada pertemuan berbagai bidang ilmu seperti machine learning, statistik, pengenalan pola, pemrosesan bahasa alami, teknologi basis data, visualisasi, interaksi komputer-manusia (HCI), algoritma, komputasi berkinerja tinggi, ilmu sosial, dan berbagai domain aplikasi lainnya [10]. Secara spesifik, dalam data mining terdapat metode khusus yang didasarkan pada tujuan pemanfaatan kumpulan data, termasuk dalam hal estimasi, prediksi, klasifikasi, klusterisasi, dan asosiasi [11].

2.6. Machine Learning

Machine Learning merupakan cabang *Artificial Intelligent* dengan mengembangkan algoritma agar memungkinkan komputer belajar berdasarkan data, sehingga sering disebut sebagai pembelajaran data. Oleh karena itu, *Machine Learning* dapat dianggap sebagai proses pemrograman komputer yang menggunakan data historis untuk membentuk model dan memungkinkan pembelajaran dengan tujuan mendapatkan performa yang optimal dalam menggali informasi dari suatu set data. Inti dari *Machine Learning* adalah menciptakan model yang dapat merepresentasikan pola-pola dalam kumpulan dataset [12].

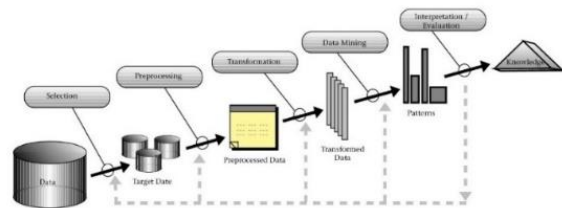
2.7. RapidMiner

RapidMiner merupakan perangkat lunak sumber terbuka (*open source*) yang juga berfungsi sebagai solusi untuk melakukan analisis dalam proses data mining, *text mining* dan analisis prediksi. *Software* ini menggunakan berbagai teknik prediksi dan deskriptif

untuk memberikan wawasan atau pengetahuan kepada pengguna, sehingga dapat membuat keputusan yang optimal. *Software* ini dapat berdiri sendiri sebagai alat analisis data dan juga dapat diintegrasikan sebagai mesin data mining pada produk lainnya [13].

3. METODE PENELITIAN

Tahapan perancangan disusun berdasarkan proses *Knowledge Discovery In Database* (KDD) bertujuan untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma.



Gambar 2. Proses KDD

Adapun penjelasan dari gambar 2 tentang tahapan proses KDD diatas, yaitu sebagai berikut:

1. Data Selection

Merupakan tahap dari KDD dalam melakukan proses pemilihan data dari suatu kumpulan data, tahap ini harus dilaksanakan sebelum memulai pencarian informasi. Data yang telah dipilih kemudian digunakan untuk proses data, disimpan dalam file, dan dipisahkan dari basis data operasional.

2. Preprocessing atau Cleaning

Sebelum melibatkan diri dalam proses data mining, penting untuk melakukan proses pembersihan pada data yang menjadi fokus *Knowledge Data Discovery*. Proses pembersihan data melibatkan langkah-langkah seperti menghapus data yang identik atau data duplikat, memeriksa konsistensi data, dan memperbaiki kesalahan pada data, termasuk kesalahan cetak. Selain itu, terdapat proses pengayaan data, dengan melibatkan penambahan data atau informasi yang relevan atau sesuai dengan data yang ada.

3. Transformation

Merupakan tahapan transformasi sehingga data yang akan digunakan menjadi sesuai untuk menjalankan proses data mining. Proses ini bergantung pada jenis informasi yang diinginkan dalam suatu basis data.

4. Data mining

Data mining merupakan proses mencari pola, hubungan, atau informasi menarik dalam data dengan memanfaatkan teknik atau metode khusus. Pemilihan metode dan algoritma ini harus tepat, dan sangat bergantung kepada tujuan keseluruhan dari proses *Knowledge Data Discovery* (KDD).

5. Interpretation atau evaluation

Merupakan representasi pola dari informasi yang dihasilkan dari proses data mining, perlu disajikan dalam format yang dapat dimengerti dengan

mudah oleh pihak yang berkepentingan. Tahap ini dikenal sebagai *interpretation* merupakan bagian integral dari proses *Knowledge Data Discovery* (KDD). Pada tahap ini, dilakukan pemeriksaan untuk memastikan apakah pola informasi yang ditemukan sesuai atau bertentangan dengan fakta atau hipotesis yang ada pada sebelumnya [14].

4. HASIL DAN PEMBAHASAN

4.1. Data selection

Dalam tahap data *selection*, data yang digunakan dalam tugas akhir yaitu data historis saham PT Bank

Mandiri selama 5 tahun dari tahun 2018 sampai 2023 dengan jumlah baris data 1255. Pada tahap ini penulis menyeleksi atribut-atribut yang akan digunakan yaitu *date*, *open*, *high*, *low*, *close*, dan *volume*. Pada tahap ini terlebih dahulu mengambil data di website *finance.yahoo.com* kemudian import data kedalam *rapidminer*. Setelah itu gunakan operator *select attributes* untuk menyeleksi atribut yang akan digunakan.

Tabel 2. Atribut yang diseleksi beserta tipe datanya

No	Date	Open	High	Low	Close	Volume
Tipe data	<i>date</i>	<i>real</i>	<i>real</i>	<i>real</i>	<i>real</i>	<i>integer</i>
1	01/11/2018	3.450	3.675	3.438	3.613	185151800
2	02/11/2018	3.663	3.750	3.550	3.713	143785200
3	05/11/2018	3.700	3.700	3.625	3.688	56013400
4	06/11/2018	3.738	3.738	3.625	3.675	124000800
5	07/11/2018	3.688	3.700	3.600	3.675	92281000
6	08/11/2018	3.700	3.763	3.688	3.750	130444200
7	09/11/2018	3.625	3.663	3.575	3.613	123465600
8	12/11/2018	3.613	3.613	3.488	3.488	65689800
9	13/11/2018	3.400	3.588	3.400	3.575	85750000
10	14/11/2018	3.638	3.650	3.563	3.563	68203400
11	15/11/2018	3.625	3.725	3.575	3.700	115193000
12	16/11/2018	3.738	3.813	3.688	3.725	113146000
...
1254	28/11/2023	5.900	5.925	5.875	5.900	58255800
1255	29/11/2023	5.850	5.925	5.825	5.850	36796600

Tabel 2 diatas merupakan tampilan data historis saham dengan atribut yang digunakan beserta tipe datanya. Atribut *date*, *open*, *high*, *low*, dan *volume* sebagai atribut independen. Dan atribut *close* sebagai label nya atau atribut dependen.

4.2. Preprocessing atau cleaning data

Pada tahap *pre-processing/cleaning* data, dilakukan proses penghapusan data jika terdapat data yang *missing value*. Lakukan pengecekan terlebih dahulu, apakah dataset yang akan di proses terdapat *missing value* atau tidak. Dapat dilakukan dengan cara memanfaatkan operator *Replace Missing Values*. Jika operator *Replace Missing Values* telah dijalankan, hasilnya dapat dilihat pada bagian *result* RapidMiner akan terlihat pada kolom *missing* di setiap atribut nya, jika nilai nya 0 menandakan bahwa tidak adanya *missing value* pada data. Oleh karenanya, proses bisa dilanjutkan ke langkah selanjutnya.

4.3. Transformation

Pada tahap ini, terjadi modifikasi terhadap struktur data yang belum memenuhi format yang sesuai dengan algoritma yang akan digunakan. Teknik yang dapat digunakan dalam persiapan data melalui tahap normalisasi. Normalisasi bertujuan untuk mengubah nilai-nilai dari suatu fitur ke dalam skala yang seragam, dan dapat dilakukan menggunakan operator *normalization* pada RapidMiner. Lakukan

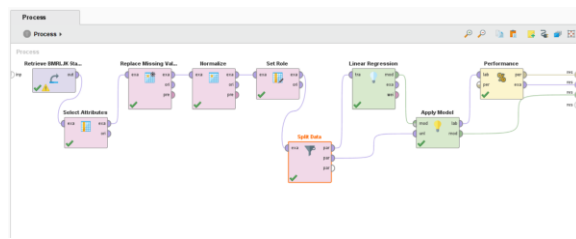
setting parameters terlebih dahulu dengan memilih *method range transformation*. Dimana nilai-nilai akan dipetakan ke dalam skala 0 sampai 1.

Tahap selanjutnya dalam *transformation* yaitu dengan mengubah *role* dari fitur tertentu untuk dijadikan label dalam dataset. Pada tahap ini, penulis akan mengubah (*change role*) atribut *close* untuk dijadikan label pada dataset tersebut dengan memanfaatkan operator *Set Role*. Lakukan *setting parameters* untuk baris *attribute name* di isi dengan atribut yang dijadikan label yaitu *close*, selanjutnya pada baris *target role* pilih role label untuk menandakan bahwa atribut *close* telah disetting menjadi label pada dataset tersebut.

Selanjutnya lakukan pembagian data dengan menggunakan operator *split data* untuk membagi dataset menjadi data *training* atau data latih dan data *testing* atau data uji. Data latih akan digunakan oleh algoritma regresi linier dalam membangun model, sedangkan data *testing* akan menjadi pengujian dalam membandingkan prediksi model yang sudah dibuat. Lakukan *setting parameters* untuk bagian *partitions* klik *edit enumeration*, selanjutnya lakukan pembagian rasion 9:1, sehingga sekitar 90% dari data atau sebanyak 1130 akan dijadikan sebagai data *training*.

dan 10% atau sebanyak 125 data sebagai data *testing*.

4.4. Data Mining



Gambar 3. Model rapidminer dengan regresi linier

Pada gambar 3 dapat dilihat rangkain proses dan operator yang digunakan dalam rapidminer menggunakan algoritma *Linear Regression*. Pada tahap ini, tambahkan operator *Linear Regression*, *apply model* dan *performances (regression)*. Pada operator *performances* lakukan *setting parameters* dengan memilih informasi apa saja yang akan ditampilkan pada hasil *performance* dari model yang telah dibangun. Pada tahap ini, penulis melakukan ceklist pada *root mean squares error*, *absolute error*, *relative error*, *correlation*, dan *squared correlation* agar nilai yang diinginkan dapat muncul pada bagian *performanceVector* saat model dijalankan.

4.5. Interpretation atau evaluation

Berikut hasil grafik dari model yang telah dibangun.



Gambar 4. Grafik harga penutupan saham

Berdasarkan gambar 4, gambar tersebut merupakan grafik pergerakan harga penutupan saham. Dapat dilihat bahwa grafik yang berwarna hijau merupakan harga penutupan yang sebenarnya. Sedangkan untuk grafik yang berwarna biru merupakan hasil prediksi harga saham penutupannya. Dari grafik tersebut terlihat pergerakan antara harga saham dan harga prediksi saham tidak jauh berbeda karena menunjukkan fluktuasi harga saham yang relatif sama.

Berdasarkan gambar 5, merupakan hasil data testing atau data uji sebanyak 125 data. Dapat dilihat seperti pada contoh tanggal 06 november 2023 harga penutupan asli yaitu 0.486 dan harga prediksi nya 0.482 memiliki selisih 0.004 serta pada tanggal 13 november 2023 harga penutupan asli 0.474 dan harga prediksi nya 0.475 memiliki selisih 0.001. Perubahan nilai-nilai dari setiap variabel dikarenakan proses normalisasi dengan menggunakan metode *range*

transformation, sehingga nilai-nilai nya dipetakan menjadi skala 0 sampai 1.

Row No.	Close	prediction(Close)	Open	High	Low	Volume	Date
113	0.393	0.391	0.383	0.382	0.387	0.135	Jun 16, 2023
114	0.411	0.412	0.404	0.409	0.407	0.161	Jul 12, 2023
115	0.423	0.424	0.416	0.415	0.419	0.159	Jul 14, 2023
116	0.450	0.450	0.436	0.447	0.446	0.112	Jul 27, 2023
117	0.459	0.454	0.442	0.453	0.449	0.173	Aug 2, 2023
118	0.495	0.488	0.481	0.485	0.485	0.083	Aug 28, 2023
119	0.495	0.497	0.492	0.491	0.494	0.079	Sep 21, 2023
120	0.495	0.489	0.478	0.488	0.485	0.097	Sep 27, 2023
121	0.482	0.456	0.451	0.450	0.452	0.106	Oct 30, 2023
122	0.483	0.469	0.457	0.474	0.464	0.150	Nov 2, 2023
123	0.474	0.479	0.478	0.479	0.476	0.069	Nov 3, 2023
124	0.486	0.482	0.475	0.474	0.479	0.076	Nov 6, 2023
125	0.474	0.475	0.469	0.465	0.473	0.023	Nov 13, 2023

ExampleSet (125 examples, 2 special attributes, 5 regular attributes)

Gambar 5. Tabel harga asli penutupan saham dan harga prediksinya

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	1.Stat	p-Value	Code
Low	1.007	0.001	1.001	0.977	772.808	0	----
Volume	0.021	0.002	0.015	0.977	11.753	0	----
(Intercept)	-0.002	0.000	?	?	-3.701	0.000	----

Gambar 6. Tabel hasil algoritma regresi linier

Berdasarkan gambar 6, koefisien atribut dan informasi terkait yaitu, variabel independen *low* dengan koefisien 1.007 menunjukkan bahwa setiap peningkatan dalam harga terendah (*low*), harga penutupan saham cenderung naik sebesar 1.007. Variabel independen *volume* dengan koefisien 0.021 menunjukkan bahwa setiap peningkatan volume, harga penutupan saham cenderung naik sebesar 0.021. *Intercept* dengan nilai -0.002 menunjukkan harga dasar (*intercept*) yang digunakan dalam model regresi.

LinearRegression

$$1.007 * Low + 0.021 * Volume - 0.002$$

Gambar 7. Hasil formula regresi linier

Formula regresi yang dihasilkan dari gambar 7, jika dituliskan dalam bentuk matematik yaitu :

$$Y = (1.007 * X_1) + (0.021 * X_2) - 0.002 \quad (7)$$

Dimana:

Y = variabel yang akan diprediksi

1.007 = nilai koefisien regresi yang dihasilkan dari proses *modelling* di RapidMiner untuk variabel *low*

0.021 = nilai koefisien regresi yang dihasilkan dari proses *modelling* di RapidMiner untuk variabel *volume*

-0.002 = *intercept* yang dihasilkan dari model RapidMiner

X_1 = variabel bebas *low*

X_2 = variabel bebas *volume*

Karena koefisien *volume* cenderung rendah maka dalam perhitungan prediksi untuk nilai baru, hanya akan memanfaatkan hasil koefisien pada variabel *low*. Data yang diambil dalam membuat model prediksi

harga saham hanya sampai tanggal 29 november 2023. Oleh karena itu, dalam melakukan prediksi harga saham di tanggal 30 november 2023 dengan estimasi harga terendah 5825 sebagai variabel x nya, maka harga penutupan dihari tersebut dapat diprediksi menggunakan rumus dibawah:

$$\begin{aligned}\text{Harga penutupan} &= (1.007 * 5825) - 0.002 \\ \text{Harga penutupan} &= 5865.775 - 0.002 \\ \text{Harga penutupan} &= 5865.773\end{aligned}\quad (8)$$

Dari perhitungan diatas dapat disimpulkan, berdasarkan estimasi harga terendah pada tanggal 30 november 2023 sebesar 5825 maka setelah dilakukan perhitungan menggunakan formula regresi, menghasilkan prediksi harga penutupan saham di harga 5865.773

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 0.005 +/- 0.000
absolute_error: 0.004 +/- 0.003
relative_error: 2.04% +/- 3.49%
correlation: 0.999
squared_correlation: 0.998
```

Gambar 8. Hasil *performanceVector*

Berdasarkan hasil *performances* model regresi linier seperti pada gambar 8, diantara temuan yang dapat diambil adalah hasilnya menunjukkan bahwa model regresi memiliki tingkat kesalahan yang rendah. Hal ini dapat dilihat dari nilai *Root Mean Squared Error* (RMSE) sebesar 0.005 +/- 0.000, yang menunjukkan seberapa dekat prediksi model dengan nilai sebenarnya.

Selain itu, *Absolute Error* memiliki nilai 0.004 +/- 0.003, yang menggambarkan selisih antara nilai prediksi dan nilai sebenarnya. Semakin kecil nilai RMSE dan *Absolute Error*, semakin baik kualitas prediksi model. Selain itu, tingkat korelasi antara variabel dalam model regresi, hasilnya menunjukkan tingkat korelasi yang kuat antara variabel X dan variabel Y , dengan nilai korelasi sebesar 0.999. Hal ini mengindikasikan hubungan yang positif dan signifikan antara variabel-variabel tersebut dalam memprediksi harga saham. Kemudian, *Squared Correlation* memiliki nilai 0.998 yang menggambarkan presentase variasi dalam variabel Y yang dapat dijelaskan oleh variabel X .

Secara keseluruhan hasil analisis menunjukkan bahwa model regresi linier yang digunakan memberikan hasil yang baik dalam memprediksi harga saham.

5. KESIMPULAN DAN SARAN

Berdasarkan proses yang telah dilakukan, dapat disimpulkan bahwa hasil prediksi menggunakan regresi linier untuk harga penutupan saham dihari selanjutnya yaitu pada tanggal 30 november 2023 dengan asumsi harga terendah 5825 seperti pada rumus8 menyentuh harga penutupan sebesar

5865.773. Untuk prediksi harga saham dihari-hari selanjutnyadapat dilakukan perhitungan kembali menggunakan rumus regresi linier yang didapat dari proses *modelling* dengan menentukan terlebih dahulu estimasi harga terendah saham sebagai variable x nya. Berdasarkan hasil *performanceVector* pada RapidMiner dengan model prediksi saham menggunakan algoritma regresi linier menghasilkan *RMSE*: 0.005 +/- 0.000, *absolute error*: 0.004 +/- 0.003, *relative error*: 2.04% +/- 3.49%, *correlation*: 0.999, dan *squared correlation*: 0.998 menunjukkan hasil yang baik. Saran pada penelitian selanjutnya diharapkan dapat membandingkan dengan metode algoritma *Support Vector Regression* atau *Neural Network* sehingga dapat menghasilkan model prediksi yang lebih optimal serta hasilnya dapat dibandingkan.

DAFTAR PUSTAKA

- [1] A. M. Hidayat and O. Romli, "PENGARUH RETURN ON EQUITY DAN CURRENT RATIO TERHADAP HARGA SAHAM PADA PERUSAHAAN PERBANKAN KONVENSIIONAL YANG TERDAFTAR DI BURSA EFEK INDONESIA PERIODE 2016-2020," *Jurnal Fakultas Ilmu Keislaman*, vol. 3, no. 2, 2022.
- [2] R. Febrilia, T. Wulandari, and D. Anubhakti, "IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM MEMREDIKSI HARGA SAHAM PT. GARUDA INDONESIA TBK," 2021. [Online]. Available: [http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index](http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexRatihFebriliaTriWulandari)
- [3] F. Andreas and U. Enri, "PERBANDINGAN ALGORITMA LINEAR REGRESSION, NEURAL NETWORK, DEEP LEARNING, DAN K-NEAREST NEIGHBOR (K-NN) UNTUK PREDIKSI HARGA BITCOIN," *JSI : Jurnal Sistem Informasi (E-Journal)*, vol. 14, no. 1, 2022, [Online]. Available: <http://ejournal.unsri.ac.id/index.php/jsi/index>
- [4] R. Julian and M. R. Pribadi, "Peramalan Harga Saham Pertambangan Pada Bursa Efek Indonesia (BEI) Menggunakan Long Short Term Memory (LSTM)," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 3, 2021, [Online]. Available: <http://jurnal.mdp.ac.id>
- [5] T. Athanasius, *Semua Bisa Investasi Cara Mudah Memahami Saham*, Edisi Digital. Jakarta: PT Elex Media Komputindo, 2022.
- [6] D. I. Putri and M. Y. Putra, "KOMPARASI ALGORITMA DALAM MEMREDIKSI PERUBAHAN HARGA SAHAM GOTO MENGGUNAKAN RAPIDMINER," vol. 11, no. 1, 2023, [Online]. Available: <https://www.kaggle.com/>,
- [7] J. Syahputra, R. Dias Ramadhani, and A. Burhanudin, "Prediksi Harga Saham Bank Bri

- Menggunakan Algoritma Linear Regresion Sebagai Strategi Jual Beli Saham,” *Data Institut Teknologi Telkom Purwokerto*, vol. 2, no. 1, pp. 1–10, 2022, [Online]. Available: <http://journal.ittelkom-pwt.ac.id/index.php/dinda>
- [8] Merfin and R. S. Oetama, “Prediksi Prospek Harga Saham Perusahaan Perbankan Menggunakan Regresi Linear (Studi Kasus Bank BCA Tahun 2015-2017),” *JSI: Jurnal Sistem Informasi (E-Journal)*, vol. 11, no. 1, 2019, [Online]. Available: <https://ejournal.unsri.ac.id/index.php/jsi/index>
- [9] Zahra, “Analisis Prediksi Jumlah Penerimaan Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana,” *Bulletin of Applied Industrial Engineering Theory*, vol. 3, 2022.
- [10] I. D. Sudirman, *Data-Driven Entrepreneur Bisnis berdaya saing dengan Data Science dan Rapidminer*. Jakarta: Salemba Infotek, 2023.
- [11] P. M. S. Tarigan, Jaya Tata Hardinata, Hendry Qurniawan, M Safii, and Riki Winanjaya, “Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus: Toko Sinar Harahap),” *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 2, pp. 9–19, Apr. 2022.
- [12] A. D. Sidik, A. Ansawarman, K. Kunci, J. Kendaraan Bermotor, M. Regresi, and F. Jalan, “Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning,” *Formosa Journal of Multidisciplinary Research (FJMR)*, vol. 1, no. 3, pp. 559–568, 2022, doi: 10.55927.
- [13] D. A. C, Donny Aji Baskoro, Lia Ambarawati, and I Wayan Simri Wicaksana, *Belajar Data Mining dengan Rapid Miner*. Jakarta, 2013.
- [14] R. Bagus Bambang Sumantri and E. Utami, “Penentuan Status Tahapan Keluarga Sejahtera Kecamatan Sidareja Menggunakan Teknik Data Mining,” *Jurnal Teknologi Informasi*, vol. 3, 2020.